Loan Defaults Prediction

Nam Tran

# Executive Summary

In the heavily regulated financial industry, developing accurate loan default prediction models is paramount for maintaining financial stability and minimizing risk. This project focuses on the development and rigorous comparison of various machine learning models designed to predict loan default probability. A key emphasis is placed on model explainability. This allows us to understand the reasoning behind predictions, which is essential for regulatory compliance and ethical decision-making. The project meticulously documents all methodologies and results, ensuring transparency and reproducibility within the regulated environment. The goal is to identify the optimal model that balances predictive power with the ability to provide clear justifications for its outcomes using proven analytical methods.

## Analysis

*Overall Observations:*

* **FICO Scores:** The median FICO score is slightly higher for non-default loans than default loans, indicating a correlation between higher scores and lower default risk.
* **Last Payment Amount:** Non-default loans tend to have higher last payment amounts than default loans. This is evident in the upward-trending box plot distribution and outliers.
* **Loan Terms** (36 months vs 60 months): Default loans are more frequently associated with longer terms compared to non-default loans. Approximately 40% of default loans have longer terms versus only 23% of non-default loans.
* **Loan Grades**: A higher proportion of default loans (25%) fall into lower credit grades (G, F, and E) compared to non-default loans (10%).
* **Feature Correlations**: Most features in the dataset exhibit low to moderate correlations. There are some highly correlated features due to their shared characteristics. These will be removed later to maintain model simplicity.

## Model Building and Evaluation

We're comparing models by looking at ROC-AUC scores on the validation dataset. A higher ROC-AUC means the model is better at correctly identifying both defaults and non-defaults. We're also paying attention to F-scores to get a balanced view of how well the model finds true defaults without too many false alarms.

*Some key results derived from model’s outputs (see figure 1):*

* Superior Performance: The fine-tuned XGBoost model outperforms other models on the validation dataset, achieving the highest ROC-AUC and F1 scores. This demonstrates its ability to accurately predict loan defaults on unseen data.
* Balanced Metrics: At a threshold of 0.5, the fine-tuned XGBoost model demonstrates strong overall performance. This is evident in the following metrics:
* Accuracy: 93%
* Recall: 78% (ability to identify default loans)
* Precision: 70% (accuracy in predicting defaults)

*5% False Positive Rate Strategy (see figure 2):*

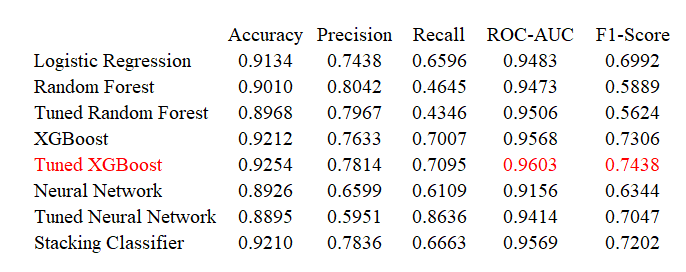
* In order to achieve the 5% False Positive Rate, we have to lower the threshold to 0.4.
* A 5% False Positive Rate means the model might mistakenly flag some good loans as bad (5% of the time), but it will also catch more actual defaults we wouldn't want to miss (79% recall).
* The model is 74% accurate in its "red flags" (74% precision). Out of every 100 loans it identifies as potentially risky, 74 actually are likely to default. However, there's a 26% chance it might mistakenly flag a good loan.
* There's a trade-off if we want to increase or decrease this threshold. We can catch more potential problems (defaults) but might waste some time investigating good loans that were flagged incorrectly. With a threshold set relatively low at 0.4, more cases are classified as potential defaults, leading to increased true positives but also more false positives.

## Recommendations

* Promote consistent, substantial payments: Customers demonstrating this behavior are less likely to default. Emphasize the importance of these payments and offer incentives or reminders for automated payment setup.
* Focus outreach on at-risk customers: Target support programs towards customers exhibiting high-risk factors such as recent credit activity, high loan amounts, high interest rates, and multiple credit lines.
* Review lending policies: Consider adjusting interest rates or loan limits based on factors like existing credit lines to tailor policies towards risk mitigation.
* Recommendation to address false negative and false positive:
* False Negative: Currently, high last payments are considered indicative of lower default risk, leading the model to misclassify some default loans as good loans. We should implement a system that spots unusual patterns in the last payment amounts made by borrowers. If someone makes an exceptionally large payment compared to what's typical for someone at risk of defaulting on their loan, the system will flag it for further review.
* False Positive: upon reviewing the false positive breakdown, it appears that recent credit activity and low payment amount mislead model to predict some good loans as default. To address this issue, the bank can look further into credit activity to payment ratio. This metric will combine recent credit activity with the size of a borrower's last payment. We should focus on cases with high ratios, which is high recent credit activity combined with a low payment, as these may be strong indicators of a higher default risk.

## Appendix

*Figure 1: Evaluation Metrics Table*



*Figure 2: ROC curve and Precision-Recall curve for False Positive Rate at 5%*

